

Chatbot-based assessment of employees' mental health: design process and pilot implementation

Article (Published Version)

Hungerbuehler, Ines, Daley, Kate, Cavanagh, Kate, Claro, Heloísa Garcia and Kapps, Michael (2021) Chatbot-based assessment of employees' mental health: design process and pilot implementation. JMIR Formative Research, 5 (4). a21678 1-11. ISSN 2561-326X

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/98854/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Original Paper

Chatbot-Based Assessment of Employees' Mental Health: Design Process and Pilot Implementation

Ines Hungerbuehler¹, PhD; Kate Daley¹, BSc, MSc, DClinPsych; Kate Cavanagh², BA, DPhil, DClinPsych; Heloísa Garcia Claro^{3,4}, BSN, MSc, PhD; Michael Kapps¹, BA

¹Vitalk, TNH Health, São Paulo, Brazil

²School of Psychology, University of Sussex, Brighton, United Kingdom

³School of Nursing, University of Campinas, Campinas, Brazil

⁴Department of Preventive Medicine, University of São Paulo, São Paulo, Brazil

Corresponding Author:

Kate Daley, BSc, MSc, DClinPsych

Vitalk

TNH Health

R. Pais Leme, 215 - Sala 2504

Pinheiros

São Paulo, 05424-150

Brazil

Phone: 55 11963883018

Email: drkatedaley@gmail.com

Abstract

Background: Stress, burnout, and mental health problems such as depression and anxiety are common, and can significantly impact workplaces through absenteeism and reduced productivity. To address this issue, organizations must first understand the extent of the difficulties by mapping the mental health of their workforce. Online surveys are a cost-effective and scalable approach to achieve this but typically have low response rates, in part due to a lack of interactivity. Chatbots offer one potential solution, enhancing engagement through simulated natural human conversation and use of interactive features.

Objective: The aim of this study was to explore if a text-based chatbot is a feasible approach to engage and motivate employees to complete a workplace mental health assessment. This paper describes the design process and results of a pilot implementation.

Methods: A fully automated chatbot (“Viki”) was developed to evaluate employee risks of suffering from depression, anxiety, stress, insomnia, burnout, and work-related stress. Viki uses a conversation style and gamification features to enhance engagement. A cross-sectional analysis was performed to gain first insights of a pilot implementation within a small to medium-sized enterprise (120 employees).

Results: The response rate was 64.2% (77/120). In total, 98 employees started the assessment, 77 of whom (79%) completed it. The majority of participants scored in the mild range for anxiety (20/40, 50%) and depression (16/28, 57%), in the moderate range for stress (10/22, 46%), and at the subthreshold level for insomnia (14/20, 70%) as defined by their questionnaire scores.

Conclusions: A chatbot-based workplace mental health assessment seems to be a highly engaging and effective way to collect anonymized mental health data among employees with response rates comparable to those of face-to-face interviews.

(JMIR Form Res 2021;5(4):e21678) doi: [10.2196/21678](https://doi.org/10.2196/21678)

KEYWORDS

chatbot; conversational agent; online; digital health; mobile phone; mental health; workplace; work stress; survey; response rate

Introduction

On average, people typically spend one-third of their lives at work. Thus, the workplace is one of the key environments that can affect quality of life, and emotional and physical well-being.

Generally, work is considered to be good for mental health with involuntary joblessness being a well-recognized risk factor for mental health problems, including depression [1]. However, stressful work conditions can also contribute to the development of mental health problems [2]. Causes of common mental health problems such as depression and anxiety are complex and may

include traumatic life experiences or inherited traits but can also be a reaction to work-related stress. The Dunedin Study found that high-demand jobs were associated with the onset of depression and anxiety in people with no prior history of diagnosis or treatment for either disorder [3]. Generally, work-related stress is considered to be a consequence of the organization and management, skills and competencies of employees, and the support they receive [2].

Besides serious consequences for an individual's mental health and associated direct medical costs, stressful working conditions have indirect costs through reduced productivity due to absenteeism and presenteeism. A recent World Health Organization-led study estimated that depression and anxiety disorders cost the global economy US \$1 trillion each year in lost productivity [4]. The implementation of cost-effective and feasible interventions could therefore have a significant impact on the individual, organization, and economy [5].

To minimize the impact of workplace risk factors, adequate policies and intervention programs should be implemented. To determine the appropriate interventions, the current mental health of employees and possible sources of work-related stress must first be assessed. There are specific measures that are commonly used to evaluate different mental health conditions and work-related risk factors, such as the Patient Health Questionnaire-9 (PHQ-9) [6]; Depression, Anxiety and Stress Scale-21 (DASS-21) [7]; and Job Satisfaction Survey (JSS) [8]. Traditionally, these questionnaires are self-administered in person or over the telephone with a health professional who facilitates completion, offering encouragement or clarification as needed. Both approaches yield similar results [9]. It is also possible to complete the scales online, using a webpage or smartphone app to display the questions and collect user responses. This method has been shown to be feasible in both clinical [10] and workplace [11] settings, albeit with relatively low response rates (34%) [11]. Symptom scores reported via a smartphone app strongly correlate with those reported through traditional paper-and-pen methods, although symptoms reported via smartphone were on average 3 points higher [12].

Engagement varies widely among digital health programs and smartphone apps [13]. According to a recent review, low engagement can occur when apps are not designed with the user in mind or do not solve the problem the user cares most about [14]. A lack of interactive or engaging features can also increase the risk of survey fatigue where users become tired and do not finish the survey [14]. Guided self-help interventions have been shown to have greater adherence than nonguided interventions [15]. This suggests that human support could improve engagement; however, this limits scalability. Chatbot-driven conversational surveys could offer an alternative by automating this encouragement and interaction. Chatbots have been shown to have significantly greater engagement and higher-quality responses than typical online surveys [16,17], and the relative anonymity offered by such an approach could be of additional value to employees.

Chatbots, or conversational computer programs, simulate human conversation. Features include word-classification processes, natural language processing, and artificial intelligence in

addition to simple keyword scans and databases linking common phrases and predefined responses, which help the chatbot tailor the answers to a specific user input. Most chatbots are accessed via websites or mobile apps, or can be integrated into virtual assistants as a conversational component of a system, which can also control external devices or manage basic tasks such as emails or to-do lists. Chatbots tend to be represented by an animated character, in some cases an embodied "human" conversational agent who uses and responds to verbal and nonverbal communication such as hand gestures or body posture.

Many interactions between organizations and customers are already bot-driven, enabling companies to respond to more people at a faster and cheaper rate than if they use human customer service representatives. Besides being a cost-effective and feasible method of communication, chatbots are capable of generating a believable and dynamic dialogue. This has the potential to enhance engagement rates, using the chatbot to successfully guide and motivate users. Compared to typical online surveys, this may result in a higher level of engagement [16] and greater symptom disclosure [17]. However, to our knowledge, no system of this nature currently exists for assessing workplace mental health in Brazil.

This paper describes the chatbot design process and results of a pilot implementation in a workplace setting. The aim of this study was to explore whether a text-based chatbot is a feasible way to engage and motivate employees to complete a workplace mental health assessment so as to provide important insights for the employee and organization.

Methods

Design

This study was a pilot implementation of a chatbot-based mental health assessment performed in a real-world workplace setting, based on a cross-sectional analysis.

Sample

The sample comprised employees of an industrial plant in São Paulo, Brazil, with a total of 120 employees, who participated in the assessment between October and November 2019. These employees work at the recycling plant (n=52); in reverse logistics operations (n=40); or in the office in administration, information technology, or human resources roles (n=28).

Approval

All participants provided consent as part of the onboarding process. They agreed for their anonymized data to be used for research purposes and for the aggregated data to be shared with the organization. Data are stored securely and are password-protected.

The company's Board of Directors approved the use of the routinely collected data for analysis and publication after having been briefed about the assessments and privacy requirements.

Research based on aggregate user data with no possibility of individual identification does not require approval of the Research Ethics Committee (CEP/CONEP) in Brazil.

Chatbot Development and Testing

The chatbot was designed to assess employees' mental health in a cost-effective and engaging way. This requires a specific developing platform, clinically validated content, and an adequate visual presentation, as well as a clear purpose and a well-defined personality. A key component for reaching this goal is the user experience (UX), which is used to help connect the chatbot with users and build a shared experience. The first step was to analyze the needs, characteristics, and behaviors of the target user group. Based on these insights, the chatbot avatar

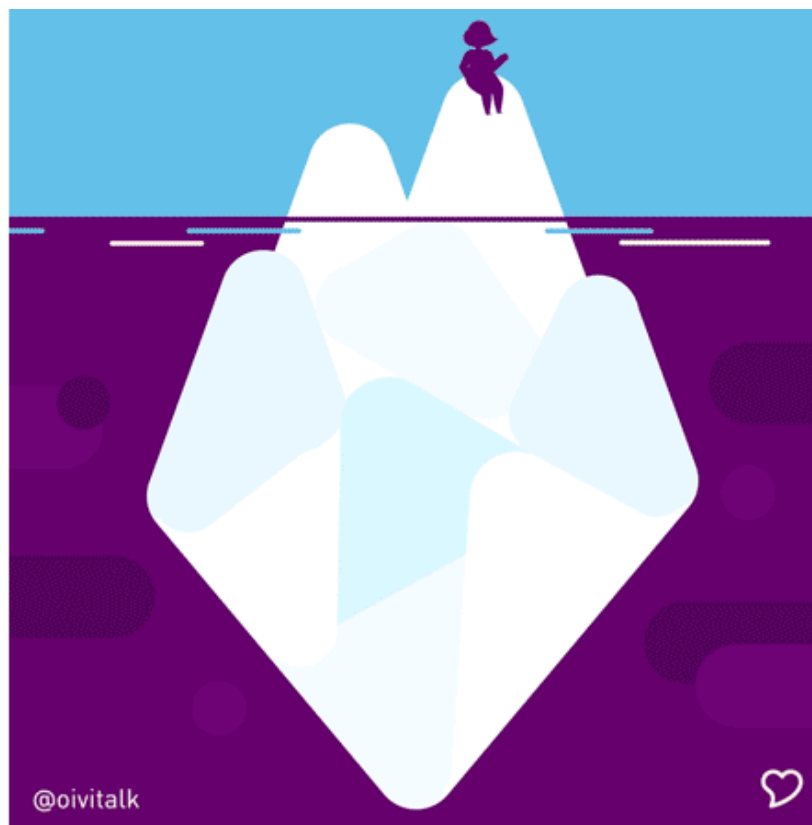
"Viki" (Figure 1) was created, along with a specific language and conversation style. The UX team performed focus groups and telephone interviews with potential users to verify if the chatbot fits with their expectations and needs. For example, they were asked if the objective was clear, if the length of the checkup was accurate, whether they had any difficulties in completing the checkup, how they rated their experience of communicating with the chatbot, and how they would describe Viki. The data obtained from these interviews were used to refine the chatbot.

Figure 1. The chatbot avatar Viki.



The chatbot is built on ruby and javascript, and was created by the team at TNH Health. The checkup assessment is rule-based, with the next steps determined by user responses. As part of the design process, decision trees were defined, and all possible user journeys were mapped and analyzed. Decision trees enabled the chatbot to provide the right responses and information based on the user's inputs (ie, to customize the conversation). For example, if the user responds that they want to know more about depression, the chatbot delivers further information about the topic. If the user prefers to continue without knowing more about the topic, the chatbot takes them to the next topic. Building a decision tree creation tool with all of the necessary settings for interactions and data organization allowed the chatbot to be updated in an agile way without needing a new system release, which was deemed to be crucial to the development.

Given known challenges in engagement, gamification features were added to address this issue. Gamification is defined as "the use of game design elements in nongame contexts" [18], with features such as levels, challenges, points, progress, feedback, story, and reward [19]. The most important features utilized in this design were story and feedback. For the story, Viki guides users through the assessment process, presented as an expedition around an iceberg (Figure 2). The iceberg represents issues relating to mental health and is divided into different sections: stress, anxiety, depression, burnout, and work-related stress. Insomnia is additionally presented if the user scores positive to the question "I had trouble falling or staying asleep" or "sleeping too much" in the depression block.

Figure 2. The iceberg story.

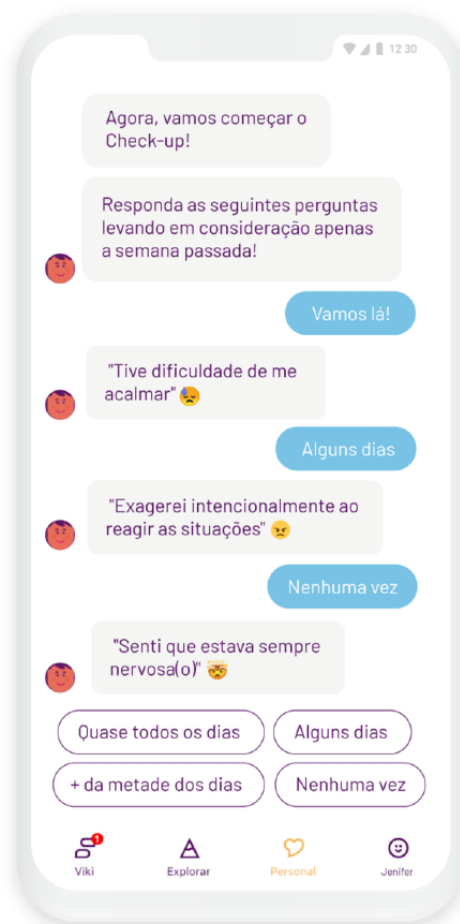
During the conversations, more and more of the iceberg, and thus the key topics, becomes visible to the user. Each section consists of an introduction to the topic and a standardized questionnaire. The questionnaires are delivered in a conversation format, with Viki asking a question and the user selecting their answers from predefined responses using the standard options for each questionnaire (Figure 3). The next question is only presented once the user selects their answer. This is displayed on a rolling screen with approximately three responses visible at a time. The user can go back and change their answer if they want, but there is no functionality included to skip a question. During the checkup, besides the questions for the different scales, Viki also offers messages of encouragement. These messages are designed to keep the participant motivated and engaged, matching the methods real human interviewers use. An example of such a message is “We are almost done, you are doing very well!” All communication is text-based, written in Brazilian Portuguese.

The entire assessment takes approximately 15 minutes. Responses are captured automatically and stored within a secure database. The user issue of nonresponse was discussed in the design process and the team agreed that a limit of 24 hours should be set. The user can pause at any time during the

assessment, but if they do not complete the assessment within 24 hours the system will reset and data are overwritten. This ensures that all questionnaire data are collected within a specific time period.

Immediately after completing the assessment, participants receive personalized feedback and recommendations. Participants are reminded that the results do not offer a diagnosis, but may indicate the presence of an emotional problem. In the case of a serious risk identified (ie, very high risk for anxiety or depression or suicide risk detected during the assessment, or via key words typed by users), the conversation follows a safety protocol that includes referral to the care network or, if necessary, to emergency services.

Aggregated and anonymized data are presented on a dashboard and in an organization report to the company’s management. This allows oversight of the assessment process in real time (eg, response rate, distributions of mental health outcome categories of already assessed employees), and offers valuable insights to the organization, identifying issues and recommending actions such as further campaign actions in case of a general low engagement of the employees and planning of target-oriented interventions for specific departments or positions based on the mental health outcome distributions.

Figure 3. Screenshot of the mental health checkup.

Participants are informed within the terms of use, which they are asked to agree to at the beginning of the checkup, about the anonymized sharing of their data with the company. To preserve anonymity, no name or other personal identification data are included on the dashboard or in the report, and departments or sections with less than 8 people were pooled into larger groups.

The usability and technical functionality of the chatbot and dashboards were tested prior to launching with users (with internal staff and organization employees). The only issue raised during this testing was how the chatbot would react to any unexpected free text entered by the employee during the assessment. Therefore, the settings and natural language understanding were adjusted to enable smooth running of the chatbot while adhering to risk protocols.

Mental Health Outcomes

The checkup includes the following questionnaires to cover the topics of anxiety, depression, stress, burnout, and work-related stress. Insomnia was added for users who did not answer the question about difficulties with sleeping with “Not at all” in the previously applied PHQ-9. All questionnaires were translated and validated in Brazilian Portuguese, with good psychometric properties [20-25]. No changes were made to question order or wording. The Generalized Anxiety Disorder-7 scale is a 7-item self-report scale used to assess anxiety symptoms over the past 2 weeks (eg, “How often have you been bothered by feeling afraid something awful might happen?”). Responses range from

0 (not at all) to 3 (nearly every day). Total scores are divided into four categories: none (0-4), mild (5-9), moderate (10-14), and severe (15+) symptoms [26]. The PHQ-9 is a 9-item self-report scale that evaluates symptoms of depression over the past 2 weeks (eg, “How often have you been bothered by feeling down, depressed, or hopeless?”). Responses range from 0 (not at all) to 3 (nearly every day). Total scores are divided into five categories: none (0-4), mild (5-9), moderate (10-14), moderately severe (15-19), and severe (20+) symptoms [6].

The DASS-21 is a 21-item self-report questionnaire consisting of three scales to measure depression, anxiety, and stress [27]. The stress subscale consists of 7 items and the user is asked how much each statement applied to them in the past week (eg, “I found it difficult to relax”). Responses range from 0 (did not apply) to 3 (applied very much or most of the time). Total scores are doubled and divided into five categories: normal (0-14), mild (15-18), moderate (19-25), severe (26-33), and extremely severe (34+) levels of stress [7].

The Insomnia Severity Index (ISI) is a 7-item self-report questionnaire assessing the nature, severity, and impact of insomnia. Responses range from 0 (no problem) to 4 (very severe problem). Total scores are categorized as: absence (0-7), subthreshold (8-14), moderate (15-21), and severe (22-28) insomnia [28].

The Oldenburg Burnout Inventory is a 16-item self-report scale with two dimensions: exhaustion and disengagement from work

(eg, “There are days when I feel tired before I arrive at work”). Answers range from 1 (strongly agree) to 4 (strongly disagree) [29]. For analysis, the following categories were used: very low (0-15), low (16-30), high (31-45), and very high (46+) risk of burnout.

The JSS questionnaire is based on a two-dimensional theoretical model created by Robert Karasek that relates two aspects at work, demands and control, to the risk of illness. A third dimension, support, was later added by Theorell [8] as a protective factor to guard against work-related stress. A 15-item short form of the questionnaire was used to assess work-related risk and protective factors in this pilot (eg, “Does your work demand too much effort? Do you have a choice in deciding how you do your work?”). Scores range from 1 (almost never/totally agree) to 4 (frequently/totally disagree). Categories used for analysis were normal (0-15), slightly increased (16-30), increased (31-45), and extremely increased (46+).

Recruitment and Implementation

As the first stage of implementation, the mental health assessment was presented to the managers of each department of the company. Subsequently, an internal multichannel information campaign was used to present the assessment to employees and educate them about the importance of mental health. Viki was introduced as part of this campaign, using a variety of online and offline channels (email, intranet, banners, and leaflets). This psychoeducative action aimed to reduce stigma and motivate employees to consider looking at their own mental health, as well as to generate trust in the product.

All employees were then emailed a weblink with an invitation to complete the checkup to obtain feedback on their scores. No financial incentive was offered, and employees were informed that the checkup was anonymous and voluntary. They were told that it would take up to 15 minutes, and no stipulations were made regarding whether to complete the checkup at work or at home. It could be completed via cell phone or computer.

When employees click the link, they first pass through an authentication process where they register and enter their company code (which helped to identify the participants of this specific company survey within the whole user population) and agree to the terms and conditions. Viki then begins the checkup.

Statistical Analysis

Descriptive statistics are used to summarize sample characteristics, baseline symptoms, and completion rates. Only complete datasets were included in this analysis, and duplicates were not possible due to the registration process used. The Checklist for Reporting Results of Internet E-surveys was used for data reporting [30].

Results

Sample and Engagement

Of the 120 eligible employees, 98 employees started the assessment (81.7% initiation rate) and 77 completed it (79% completion rate). This response rate enabled data to be obtained for 64.2% (77/120) of the organization's workforce. Data collection was performed within 4 weeks. More than half of those responding did so within the first week of the campaign.

Of the 77 employees completing the checkup, the majority were men (45/77, 58%), white (40/77, 52%), and aged between 25 and 44 years (55/77, 71%), reflecting the demographics of the organization's workforce. The majority of respondents were in occupations of plant and machine operators and assemblers (n=21, 27%) or technicians and associate professionals (n=19, 25%).

Implementation

During implementation, we noticed differences in the speed of rollout, and considered if the motivation and conviction at a management level could be a factor in engagement. As a preventative measure to maximize engagement, we contacted department managers to emphasize the benefits of the checkup and to troubleshoot any potential issues. No issues were identified, but some managers reported they would ensure that they pass on this information to their teams. The informal feedback received from the lead of the project was that the campaign had been well received and they felt that the chatbot added value. This evidence is of course more anecdotal as it did not form part of the data analysis.

Mental Health Risks

Overall, the sample scored in the low ranges on the majority of the questionnaires (see Table 1). Scores were higher in the area of work-related stress, with equal control and demand reported alongside lower support. On the PHQ-9, 44% (34/77) of users reported difficulties with sleep and therefore also completed the ISI for insomnia.

For those with symptoms, the majority of respondents scored in the mild range for anxiety (20/40, 50%), mild range for depression (16/28, 57%), subthreshold level for insomnia (14/20, 70%), and moderate level for stress (10/22, 46%) as defined by the questionnaire scores. For burnout, most of the respondents scored in the low-risk category (50/74, 68%), whereas the majority scored in the increased-risk category for job-related stress (53/77, 69%).

Table 1. Scores and outcomes for each questionnaire measure.

Measure	Respondents (n)	Mean (SD)	Assessment category
Anxiety (GAD-7 ^a)	77	6.21 (4.56)	Low
Depression (PHQ-9 ^b)	77	4.40 (5.21)	None
Stress (DASS-21 ^c)	77	11.09 (7.13)	Normal
Insomnia (ISI ^{d,e})	34	9.26 (5.66)	Subthreshold
Burnout (OLBI ^f)	77	27.68 (8.38)	Low
Occupational stress (JSS^g)	77	32.38 (3.55)	Increased
Control	77	12.32 (1.99)	Increased
Demand	77	12.19 (1.72)	Increased
Support	77	7.86 (2.40)	Increased

^aGAD-7: Generalized Anxiety Disorder-7.

^bPHQ-9: Patient Health Questionnaire-9.

^cDASS-21: Depression, Anxiety and Stress Scale-21.

^dISI: Insomnia Severity Index.

^eCompleted only if sleep was identified as an issue.

^fOLBI: Oldenburg Burnout Inventory.

^gJSS: Job Stress Scale.

Discussion

Principal Findings

The chatbot-based assessment was successfully implemented in the workplace, suggesting that a chatbot could be a feasible way to engage employees in completing a workplace mental health assessment. There was a 79% (77/98) completion rate, obtaining questionnaire responses from 64.2% (77/120) of the workforce. This compares favorably to face-to-face data collection methods. For example, the São Paulo Megacity study reported a response rate of 81% [31] and epidemiological studies reported a response rate of 70% using the same method of data collection [32]. The completion rate in this analysis is 25% higher than that of results found in a telecommunication company using a web-based screening for depression [11], and is also clearly higher than that of other online surveys where rates below 10% are common [33]. A previous study reported a similar completion rate (78%) using a smartphone app to administer up to three survey sessions per day in a sample of psychiatric outpatients. However, this involved a smaller sample and the participants were financially compensated, which may influence motivation [12]. It would be interesting to further explore the impact of the different components of implementation to ascertain the factors that are most integral for success, such as the chatbot or the onboarding process. An online tool and distribution method could be particularly pertinent in an era of remote working.

The majority of users obtained low scores on all of the questionnaires. However, many users did report symptoms at a moderate or severe level. The proportions scoring at this level for depression (20/77, 26%) and anxiety (12/77, 16%) are higher than the estimated prevalence rates of anxiety disorders (9.3%) and depressive disorders (5.8%) found in the general population

of Brazil [34], although these rates do vary considerably depending on the method of data collection and the measure used [35]. Prevalence rates are typically higher in women than in men [35], which is interesting given that this sample predominately comprised men. The risk of burnout identified is similar to levels found in other professional groups such as health workers [36] and teachers [37]. High demand was often reported in combination with a low degree of perceived control, which are both risk factors for burnout. This combination is considered to be the most critical in terms of a negative impact on individual mental health. The presence of anxiety, depression, and insomnia could also impact organizational productivity through absenteeism and presenteeism [2,11], which requires further exploration.

The fact that some people reported symptoms shows the importance of addressing mental health issues within the workplace. It would be interesting to determine whether these people have ever sought professional help or support for these difficulties. Insights gained from the assessment could be used to identify individual- and organizational-level strategies that could be implemented to improve mental health, and potentially productivity, within the workplace.

Limitations

The questionnaire measures have not yet been validated for use in a chatbot format with gamification. As this pilot showed that implementation is feasible, the next step is to complete a validation study to assess the effect of using this method on the psychometric properties of the questionnaires. Establishing validity is required before conclusions can reliably be drawn regarding the mental health of the workplace.

Although the chatbot could successfully obtain data for 64% of the workforce, there are no data for those who did not participate or for those who dropped out during the assessment. It would

be interesting to compare sociodemographic characteristics and baseline symptoms between the groups. For example, previous research indicates that being male, with a lower educational level, and comorbidity of depression with anxiety can increase the risk of dropout and nonengagement [38]. Such insights could be used to adapt the chatbot or onboarding process to make it more appealing to the less-engaged group. Taking age and gender into account has been found to enhance use of digital mental health programs [39].

As this was a cross-sectional data analysis, we do not know if the high response rate will be sustained over time. There is the possibility that the response rate was inflated due to the novelty of the approach, which has been suggested to be a general characteristic of mobile health interventions [40]. Longitudinal research could explore this issue, which could assist in the development of UX with chatbots over time. Additionally, as the system used for output generation (responses to the user's input) is fixed (based on predefined decision trees), the conversation opportunities are currently limited. A larger amount of conversational data gained over time will be necessary to train the chatbot and make it intelligent in a more autonomous way. This would also allow for free text interpretation, for example.

Without controls, it is unclear if the results are due to the specific sample, their workplace, or the method of data collection, which may facilitate higher levels of disclosure [11]. After validating the questionnaires for use in this format, it would be important to repeat this exercise with different workplaces to ascertain if the results are generalizable to other populations, particularly

considering issues of gender, age, work type, and level of education. It would be important to replicate the implementation in a workplace with more gender balance and a larger sample size. A usability questionnaire would further strengthen the validity of the results.

Further research is planned, which will include comparison with traditional paper-and-pen methods and web-based forms. These studies will also include repeat measurement following the implementation of remedial measures to ascertain their impact. We would also like to explore, using implementation science, the factors and processes involved in successful implementation of the chatbot within the workplace since there are many variables involved.

Conclusions

The creation of healthy workplaces and adequate mental health policies must be based on a comprehensive needs assessment. Face-to-face assessments are not anonymous or scalable, and online surveys are often limited by low response rates. A chatbot offers a fully automated digital solution, incorporating gamification features to engage and motivate employees to complete a workplace mental health assessment. The chatbot was found to have response rates comparable to those of face-to-face interviews, suggesting that this could be a feasible way to collect such data. To further verify this new solution, a validation study comparing it with other formats such as face-to-face interviews or online surveys, as well as including a feasibility and satisfaction analysis, would be the next logical step.

Acknowledgments

The authors are grateful to Raphael Mota and Bruno de Castro Paul Schultze for assisting with the additional data extraction and analysis, and to Aimee da Silva Ferreira and Carolina Nakatsu for the graphic design. Funding for the study was provided by TNH Health.

Authors' Contributions

IH made substantial contributions to the conception and design of the work, and completed analysis of the results. All authors contributed to the writing, review, and approval of this manuscript for publication.

Conflicts of Interest

IH, KD, and MK work for TNH Health, the company that developed the chatbot and provided funding for this publication. KC and HC declare no conflicts of interest.

References

1. Jefferis BJ, Nazareth I, Marston L, Moreno-Kustner B, Bellón JÁ, Svab I, et al. Associations between unemployment and major depressive disorder: evidence from an international, prospective study (the predict cohort). *Soc Sci Med* 2011 Dec;73(11):1627-1634. [doi: [10.1016/j.socscimed.2011.09.029](https://doi.org/10.1016/j.socscimed.2011.09.029)] [Medline: [22019370](https://pubmed.ncbi.nlm.nih.gov/22019370/)]
2. Mental Health in the Workplace. World Health Organization. URL: http://www.who.int/mental_health/in_the_workplace/en/ [accessed 2020-02-11] [[WebCite Cache ID](#) www.who.int/teams/mental-health-and-substance-use/mental-health-in-the-workplace]
3. Melchior M, Caspi A, Milne BJ, Danese A, Poulton R, Moffitt TE. Work stress precipitates depression and anxiety in young, working women and men. *Psychol Med* 2007 Aug;37(8):1119-1129 [[FREE Full text](#)] [doi: [10.1017/S0033291707000414](https://doi.org/10.1017/S0033291707000414)] [Medline: [17407618](https://pubmed.ncbi.nlm.nih.gov/17407618/)]
4. Chisholm D, Sweeny K, Sheehan P, Rasmussen B, Smit F, Cuijpers P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. *Lancet Psychiatry* 2016 May;3(5):415-424. [doi: [10.1016/S2215-0366\(16\)30024-4](https://doi.org/10.1016/S2215-0366(16)30024-4)] [Medline: [27083119](https://pubmed.ncbi.nlm.nih.gov/27083119/)]

5. Bertram M, Banatvala N, Kulikov A, Belausteguigoitia I, Sandoval R, Hennis A, et al. Using economic evidence to support policy decisions to fund interventions for non-communicable diseases. *BMJ* 2019 May 20;365:l1648. [doi: [10.1136/bmj.l1648](https://doi.org/10.1136/bmj.l1648)] [Medline: [31110025](https://pubmed.ncbi.nlm.nih.gov/31110025/)]
6. Kroenke K, Spitzer R. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann* 2002 Sep 01;32(9):509-515 [FREE Full text] [doi: [10.3928/0048-5713-20020901-06](https://doi.org/10.3928/0048-5713-20020901-06)]
7. Antony M, Bieling P, Cox B, Enns M, Swinson R. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychol Assess* 1998 Jun;10(2):176-181 [FREE Full text] [doi: [10.1037/1040-3590.10.2.176](https://doi.org/10.1037/1040-3590.10.2.176)]
8. Theorell T. The demand-control-support model for studying health in relation to the work environment: an interactive model. In: Orth-Gómer K, Schneiderman N, editors. *Behavioral medicine approaches to cardiovascular disease*. Mahwah, NJ: Lawrence Erlbaum Associates; Jun 25, 1996.
9. Pinto-Meza A, Serrano-Blanco A, Peñarribia MT, Blanco E, Haro JM. Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? *J Gen Intern Med* 2005 Aug;20(8):738-742 [FREE Full text] [doi: [10.1111/j.1525-1497.2005.0144.x](https://doi.org/10.1111/j.1525-1497.2005.0144.x)] [Medline: [16050884](https://pubmed.ncbi.nlm.nih.gov/16050884/)]
10. Fann JR, Berry DL, Wolpin S, Austin-Seymour M, Bush N, Halpenny B, et al. Depression screening using the Patient Health Questionnaire-9 administered on a touch screen computer. *Psychooncol* 2009 Jan;18(1):14-22 [FREE Full text] [doi: [10.1002/pon.1368](https://doi.org/10.1002/pon.1368)] [Medline: [18457335](https://pubmed.ncbi.nlm.nih.gov/18457335/)]
11. Harvey SB, Glozier N, Henderson M, Allaway S, Litchfield P, Holland-Elliott K, et al. Depression and work performance: an ecological study using web-based screening. *Occup Med (Lond)* 2011 May;61(3):209-211. [doi: [10.1093/occmed/kqr020](https://doi.org/10.1093/occmed/kqr020)] [Medline: [21525074](https://pubmed.ncbi.nlm.nih.gov/21525074/)]
12. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a personal smartphone custom app to assess the Patient Health Questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. *JMIR Ment Health* 2015;2(1):e8 [FREE Full text] [doi: [10.2196/mental.3889](https://doi.org/10.2196/mental.3889)] [Medline: [26543914](https://pubmed.ncbi.nlm.nih.gov/26543914/)]
13. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. *J Med Internet Res* 2018 Jun 06;20(6):e199 [FREE Full text] [doi: [10.2196/jmir.9275](https://doi.org/10.2196/jmir.9275)] [Medline: [29875089](https://pubmed.ncbi.nlm.nih.gov/29875089/)]
14. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evid Based Ment Health* 2018 Aug;21(3):116-119. [doi: [10.1136/eb-2018-102891](https://doi.org/10.1136/eb-2018-102891)] [Medline: [29871870](https://pubmed.ncbi.nlm.nih.gov/29871870/)]
15. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. *Internet Interv* 2014 Oct;1(4):205-215. [doi: [10.1016/j.invent.2014.08.003](https://doi.org/10.1016/j.invent.2014.08.003)]
16. Xiao Z, Zhou MX, Liao QV, Mark G, Chi C, Chen W, et al. Tell me about yourself: using an AI-powered chatbot to conduct conversational surveys. *ACM Trans Comput Hum Interact* 2020 Jun 16;27(3):15. [doi: [10.1145/3381804](https://doi.org/10.1145/3381804)]
17. Lucas G, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Front Robot AI* 2017 Oct 12;4:1-9. [doi: [10.3389/frobt.2017.00051](https://doi.org/10.3389/frobt.2017.00051)]
18. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". New York, NY: Association for Computing Machinery; 2011 Sep Presented at: 15th International Academic MindTrek Conference: Envisioning Future Media Environments; September 2011; Tampere, Finland p. 9-15 URL: <https://dl.acm.org/doi/10.1145/2181037.2181040> [doi: [10.1145/2181037.2181040](https://doi.org/10.1145/2181037.2181040)]
19. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. *JMIR Serious Games* 2013 Dec 12;1(1):e3 [FREE Full text] [doi: [10.2196/games.3139](https://doi.org/10.2196/games.3139)] [Medline: [25658754](https://pubmed.ncbi.nlm.nih.gov/25658754/)]
20. Moreno AL, DeSousa DA, Souza AMFLP, Manfro GG, Salum GA, Koller SH, et al. Factor structure, reliability, and item parameters of the Brazilian-Portuguese version of the GAD-7 Questionnaire. *Temas Psicol* 2016;24(1):367-376. [doi: [10.9788/tp2016.1-25](https://doi.org/10.9788/tp2016.1-25)]
21. de Lima Osório F, Vilela Mendes A, Crippa J, Loureiro S. Study of the discriminative validity of the PHQ-9 and PHQ-2 in a sample of Brazilian women in the context of primary health care. *Perspect Psychiatr Care* 2009 Jul;45(3):216-227. [doi: [10.1111/j.1744-6163.2009.00224.x](https://doi.org/10.1111/j.1744-6163.2009.00224.x)] [Medline: [19566694](https://pubmed.ncbi.nlm.nih.gov/19566694/)]
22. Patias N, Machado W, Bandeira D, Dell'Aglio D. Depression Anxiety and Stress Scale (DASS-21) - Short Form: Adaptação e Validação para Adolescentes Brasileiros. *Psico-USF* 2016 Dec;21(3):459-469. [doi: [10.1590/1413-82712016210302](https://doi.org/10.1590/1413-82712016210302)]
23. Castro L. Adaptation and validation of the Insomnia Severity Index (ISI): population characteristics, normative values and associated factors. Universidade Federal de São Paulo. 2011. URL: <http://repositorio.unifesp.br/handle/11600/23193> [accessed 2020-02-18]
24. Schuster MDS, Dias VDV. Oldenburg Burnout Inventory - validation of a new way to measure Burnout in Brazil. *Cien Saude Colet* 2018 Feb;23(2):553-562 [FREE Full text] [doi: [10.1590/1413-81232018232.27952015](https://doi.org/10.1590/1413-81232018232.27952015)] [Medline: [29412413](https://pubmed.ncbi.nlm.nih.gov/29412413/)]
25. de Mello AMG, Chor D, Faerstein E, Lopes CDS, Werneck GL. Short version of the "job stress scale": a Portuguese-language adaptation. *Rev Saude Publica* 2004 Apr;38(2):164-171 [FREE Full text] [doi: [10.1590/s0034-89102004000200003](https://doi.org/10.1590/s0034-89102004000200003)] [Medline: [15122370](https://pubmed.ncbi.nlm.nih.gov/15122370/)]
26. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006 May 22;166(10):1092-1097. [doi: [10.1001/archinte.166.10.1092](https://doi.org/10.1001/archinte.166.10.1092)] [Medline: [16717171](https://pubmed.ncbi.nlm.nih.gov/16717171/)]

27. Lovibond P, Lovibond S. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav Res Ther* 1995 Mar;33(3):335-343. [doi: [10.1016/0005-7967\(94\)00075-u](https://doi.org/10.1016/0005-7967(94)00075-u)]
28. Bastien C, Vallières A, Morin C. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med* 2001 Jul;2(4):297-307. [doi: [10.1016/s1389-9457\(00\)00065-4](https://doi.org/10.1016/s1389-9457(00)00065-4)] [Medline: [11438246](https://pubmed.ncbi.nlm.nih.gov/11438246/)]
29. Halbesleben JRB, Demerouti E. The construct validity of an alternative measure of burnout: Investigating the English translation of the Oldenburg Burnout Inventory. *Work Stress* 2005 Jul;19(3):208-220. [doi: [10.1080/02678370500340728](https://doi.org/10.1080/02678370500340728)]
30. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res* 2004 Sep 29;6(3):e34 [FREE Full text] [doi: [10.2196/jmir.6.3.e34](https://doi.org/10.2196/jmir.6.3.e34)] [Medline: [15471760](https://pubmed.ncbi.nlm.nih.gov/15471760/)]
31. Viana MC, Teixeira MG, Beraldi F, Bassani IDS, Andrade LH. São Paulo Megacity Mental Health Survey - a population-based epidemiological study of psychiatric morbidity in the São Paulo metropolitan area: aims, design and field implementation. *Braz J Psychiatry* 2009 Dec;31(4):375-386 [FREE Full text] [doi: [10.1590/s1516-44462009000400016](https://doi.org/10.1590/s1516-44462009000400016)] [Medline: [20098829](https://pubmed.ncbi.nlm.nih.gov/20098829/)]
32. Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol* 2007 Sep;17(9):643-653. [doi: [10.1016/j.annepidem.2007.03.013](https://doi.org/10.1016/j.annepidem.2007.03.013)] [Medline: [17553702](https://pubmed.ncbi.nlm.nih.gov/17553702/)]
33. Hooley T, Wellens J, Madge C, Goss S. Online research methods for mental health. In: Anthony K, Mertz Nagel D, Gross S, editors. *The use of technology in mental health: applications, ethics and practice*. Springfield: Charles C Thomas; 2010:234-245.
34. Depression and other common mental disorders: Global Health Estimates Internet. World Health Organization. 2017 Apr 7. URL: http://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/ [accessed 2020-04-16]
35. Silva MT, Galvao TF, Martins SS, Pereira MG. Prevalence of depression morbidity among Brazilian adults: a systematic review and meta-analysis. *Braz J Psychiatry* 2014 Sep;36(3):262-270 [FREE Full text] [doi: [10.1590/1516-4446-2013-1294](https://doi.org/10.1590/1516-4446-2013-1294)] [Medline: [25119639](https://pubmed.ncbi.nlm.nih.gov/25119639/)]
36. Moreira DDS, Magnago RF, Sakae TM, Magajewski FRL. Prevalence of burnout syndrome in nursing staff in a large hospital in south of Brazil. *Cad Saude Publica* 2009 Jul;25(7):1559-1568. [doi: [10.1590/s0102-311x2009000700014](https://doi.org/10.1590/s0102-311x2009000700014)] [Medline: [19578577](https://pubmed.ncbi.nlm.nih.gov/19578577/)]
37. Alves PC, Oliveira ADF, Paro HBMD. Quality of life and burnout among faculty members: How much does the field of knowledge matter? *PLoS One* 2019;14(3):e0214217 [FREE Full text] [doi: [10.1371/journal.pone.0214217](https://doi.org/10.1371/journal.pone.0214217)] [Medline: [30901373](https://pubmed.ncbi.nlm.nih.gov/30901373/)]
38. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. *Psychol Med* 2015 Oct;45(13):2717-2726. [doi: [10.1017/S0033291715000665](https://doi.org/10.1017/S0033291715000665)] [Medline: [25881626](https://pubmed.ncbi.nlm.nih.gov/25881626/)]
39. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for e-mental health programs among men and women. *J Med Internet Res* 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: [10.2196/11224](https://doi.org/10.2196/11224)] [Medline: [30882361](https://pubmed.ncbi.nlm.nih.gov/30882361/)]
40. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. *JMIR Mhealth Uhealth* 2014 May 16;2(2):e24 [FREE Full text] [doi: [10.2196/mhealth.3328](https://doi.org/10.2196/mhealth.3328)] [Medline: [25099752](https://pubmed.ncbi.nlm.nih.gov/25099752/)]

Abbreviations

DASS-21: Depression, Anxiety and Stress Scale-21

ISI: Insomnia Severity Index

JSS: Job Stress Scale

PHQ-9: Patient Health Questionnaire-9

UX: user experience

Edited by G Eysenbach; submitted 21.06.20; peer-reviewed by A McLean, J Zulueta, E Sezgin, J Li; comments to author 17.07.20; revised version received 10.09.20; accepted 17.03.21; published 21.04.21

Please cite as:

Hungerbuehler I, Daley K, Cavanagh K, Garcia Claro H, Kapps M

Chatbot-Based Assessment of Employees' Mental Health: Design Process and Pilot Implementation

JMIR Form Res 2021;5(4):e21678

URL: <https://formative.jmir.org/2021/4/e21678>

doi: [10.2196/21678](https://doi.org/10.2196/21678)

PMID:

©Ines Hungerbuehler, Kate Daley, Kate Cavanagh, Heloísa Garcia Claro, Michael Kapps. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 21.04.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://formative.jmir.org>, as well as this copyright and license information must be included.